The acyclic orientation game on random graphs

Noga Alon* Zsolt Tuza†

Dedicated to Professor Paul Erdős on the occasion of his 80^{th} birthday

Abstract

It is shown that in the random graph $G_{n,p}$ with (fixed) edge probability $p > 0$, the number of edges that have to be examined in order to identify an *acyclic* orientation is $\Theta(n \log n)$ almost surely. For unrestricted p, an upper bound of $O(n \log^3 n)$ is established. Graphs $G = (V, E)$ in which all edges have to be examined are considered, as well.

1 Introduction

In this note we investigate the typical length of the following 2-person game. Given a graph $G = (V, E)$, in each step of the game player A (Algy) selects an edge $e \in E$ and player S (Strategist) orients e in the way he likes; the only restriction is that S must not create a directed circuit. The game is over when the actually obtained partial orientation of G extends to a *unique acyclic orientation.* The goal of A is to locate such an orientation with as few questions as possible, while S aims at the opposite. Assuming that both A and S play optimally, the number of questions during the game on G is denoted by $c(G)$.

A different but equivalent formulation of this nice game was first given by Manber and Tompa [8], who were motivated by a problem of testing whether a given coloring of a graph is a proper coloring. Some recent results concerning $c(G)$ have been obtained by Aigner, Triesch and the second author in [1], including the general estimates

$$n \log \frac{n}{\alpha} - O(n) \leq c(G) \leq \alpha n (\log \frac{n}{\alpha} + 1) \tag{1}$$

where n is the number of vertices, α denotes the (vertex) independence number, and "log" means logarithm in base 2. Let us note that a related lower bound can

*Department of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel and School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540, USA. Research supported in part by a USA-Israeli BSF Grant and by the Sloan Foundation Grant No. 93-6-6.

†Computer and Automation Institute, Hungarian Academy of Sciences, H-1111 Budapest, Kende u. 13–17, Hungary. Research supported in part by the OTKA Research Fund of the Hungarian Academy of Sciences, grant no. 2569.

be deduced also from one of the results of [7] stating that a graph G with degree sequence $d_1, \ldots, d_n$ has at least $\prod_{i=1}^{n}((d_i+1)!)^{1/(d_i+1)}$ acyclic orientations. This clearly implies that for any such G,

$$c(G) \geq \sum_{i=1}^{n} \log \frac{d_i + 1}{e}.$$

Let $G_{n,p}$ denote, as usual, the random graph on n labelled vertices with edge probability p. (See, e.g., [6] for the model and some of its properties.) When the edge probability p is *fixed*, the above inequalities determine $c(G_{n,p})$ within the accuracy of a multiplicative factor of $O(\log n)$ (with probability that tends to 1 as n tends to infinity). In the present note we first derive a more exact estimate, showing that in fact $O(n \log n)$ is the correct order of magnitude, i.e., the *lower* bound is tight for all (fixed) $p > 0$.

Theorem 1 *For any fixed edge probability $p > 0$, the random graph $G = G_{n,p}$ has $c(G) = \Theta(n \log n)$ with probability $1 - o(1)$.*

Our argument proving the above theorem supplies very little information for the case where $p(n)$ tends to zero as n gets large, and it remains an open problem to analyze the exact behavior of $c(G_{n,p})$ where the edge probability $p = p(n)$ tends to zero as $n \to \infty$. It may be true, however, that $c(G_{n,p}) = O(n \log n)$ holds for all p. By (1), this bound, if true, would be tight for $p = cn^{-c'}$ for all admissible choices of the constants $c > 0$ and $0 \leq c' < 1$. Note that the gap between the upper and lower bounds in (1) increases when $p(n)$ decreases, and is a power of n when $1/p(n)$ is a power of n.

The next theorem supplies a much sharper estimate for sparse random graphs.

Theorem 2 *For any edge probability $p = p(n)$, the random graph $G = G_{n,p}$ has $c(G) = O(n \log^3 n)$ with probability $1 - o(1)$.*

The proofs of Theorems 1 and 2 are different, but both combine some of the techniques used in the study of parallel comparison algorithms (see [3], [4], [2], [9]) with several new ideas. We note that the exponent of $\log n$ in Theorem 2 can be reduced slightly below 3 at the cost of making the argument somewhat more complicated, but—as this would not reduce the exponent to less than 2, and as we suspect that the optimum value of the exponent is 1 actually—we do not present the more complicated proof.

Let us recall from [1] that another challenging unsolved problem is to prove that $c(G) \leq \frac{1}{4}n^2 + o(n^2)$ for all graphs G on n vertices. If valid, this upper bound would be best possible in general. We also note that there is no known sequence $(G_n)_{n>0}$ of graphs, where G_n has n vertices, for which the difference $c(G_n) - \frac{1}{4}n^2$ tends to infinity with n.

The proofs of Theorems 1 and 2 are presented in Sections 2 and 3, respectively. The final Section 4 contains some comments on graphs $G = (V, E)$ for which $c(G) = |E|$.

2 Fixed edge probability

In this section we prove Theorem 1. For simplicity, we denote $G_{n,p}$ by G, where p is any fixed positive edge probability. The argument is based on the following properties that hold for G almost surely. (Here and in what follows, "almost surely" always means "with probability that tends to 1 as n tends to infinity". In addition, as usual, for two positive real functions $f(x)$ and $g(x)$, the notation $f(x) = \Theta(g(x))$ means "$f(x) = O(g(x))$ and $g(x) = O(f(x))$".)

1. For some function $k = k(n)$ of order $\Theta(\log n)$, any two disjoint sets of k vertices each are joined by at least one edge.

2. There is a function $k' = k'(n) = \Theta(\log n)$ such that, for any two disjoint sets X and Y of k' vertices each, there is a vertex $x \in X$ with at least k neighbors in Y, where $k = k(n)$ is a function satisfying the requirements of (1) above.

The first property is well-known, and the second one is a fairly simple consequence of the Chernoff inequality. Indeed,the expected number of edges between X and Y is $pc^2 \log^2 n$ for $k' = c \log n$, while the nonexistence of $x \in X$ with sufficiently many neighbors in Y would admit no more than $kc \log n$ edges; and the pair X, Y can be chosen in at most $\exp(2c \log^2 n)$ different ways. Thus, choosing c sufficiently large (here "large" also depends on the value of the edge probability p) the requirement holds for all X and Y almost surely.

An essential step in the proof of Theorem 1 is the following "deterministic" statement concerning linear extensions of partial orders. To fix the notation, for an oriented *acyclic* digraph $D = (V, A)$ we denote by D^* the transitive closure of D, i.e., $D^* = (V, A^*)$ is the smallest digraph in which $A \subset A^*$ and $xy, yz \in A^*$ implies $xz \in A^*$ for all $x, y, z \in V$. Two vertices $x, y \in V$ are *comparable* if $xy \in A^*$ or $yx \in A^*$; for $xy \in A^*$ we also say "x is smaller than y" or "y is larger than x". A *linear extension* L of D is an ordering $v_1 v_2 \ldots v_n$ of V such that $i < j$ holds whenever v_i is smaller than v_j.

In the next assertion we need *not* assume that the values of k and k' are proportional to $\log n$.

Lemma 3 *Suppose that the underlying undirected graph of an acyclic oriented graph $D = (V, A)$ of order n satisfies the properties* (1) *and* (2) *above, for some k and k'. Then, in every linear extension $L = v_1 v_2 \ldots v_n$ of D, for every integer r between* 1 *and n there is a subscript i ($r - 2k' < i < r + 2k'$) for which there are at least $r - 2k'$ vertices smaller than v_i and at least $n - r - 2k'$ vertices larger than v_i.*

Proof. Consider the set $Y^+ = \{v_i \mid r + k' < i \le r + 2k'\}$. By (2), there are fewer than k' vertices in $\{v_j \mid 1 \le i \le r + k'\}$ having at most $k - 1$ neighbors in Y^+. Denote by Z^+ the set of vertices v_j having at least k neighbors in Y^+, with $j \le r + k'$. By (2), $|Z^+| > r$ holds. For each $v_j \in Z^+$, the (at least) k neighbors of v_j in Y^+ dominate all but at most $k - 1$ vertices of $\{v_i \mid r + 2k' < i \le n\}$, by (1). Thus, every vertex $v_j \in Z^+$ is smaller than at least k vertices in Y^+ and at least

$n - r - k - 2k'$ vertices following Y^+, i.e., v_j is smaller than at least $n - r - 2k'$ vertices of D. Similarly, for the set $Y^- = \{v_i \mid r - 2k' < i \le r - k'\}$ we can find a set $Z^- \subseteq \{v_j \mid r - k' < j \le n\}$ of cardinality $|Z^-| > n - r$ such that every $v_j \in Z^-$ is larger than k vertices of Y^- and $r - k - 2k'$ vertices preceding Y^-, i.e., v_j is larger than at least $r - 2k'$ vertices of D. Since $|Z^-| + |Z^+| > n$, we can choose a vertex $w \in Z^- \cap Z^+$; this $w = v_i$ satisfies the requirements of the lemma. □

We now turn to the proof of Theorem 1, locating the acyclic orientation to be found, by applying an inductive algorithm. Let v be an arbitrary vertex of the random graph $G = G_{n,p}$. Assuming that we have complete information about the orientation of $G - v$, we are going to show that the orientations of all edges incident to v can be determined by $O(\log n)$ questions (provided that G satisfies (1) and (2) above).

If the orientation of $G - v$ is not transitive, we first take its transitive closure, denoted D^*. Let $D' = (V', E')$ be the subdigraph of D^* induced by the neighbors of v. Denoting $n' = |V'|$, let $v_1 v_2 \dots v_{n'}$ be a linear extension of D'. To find the orientations of all edges from v to V', we are going to apply binary search on an appropriately chosen restricted set $V'' \subseteq V'$, and then complete the algorithm with a few further questions.

As we already know, by the properties (1) and (2), Lemma 3 implies that for every r $(1 \le r \le n')$ there is a vertex v_i which is larger than $r - 2c\log n$ vertices of V' and smaller than $n' - r - 2c\log n$ vertices of V', for some appropriately chosen constant c (we have taken $k' = c\log n$ here; note that if G satisfies (1) and (2), so does its induced subgraph on the neighbors of v). Define V'' as the set of those i satisfying the above requirements for at least one value of r. Note that the gap between any two consecutive members of V'' is smaller than $4c\log n$.

Now, by a binary search on V'' we can locate a pair $v_i, v_j \in V''$ of vertices in $\log|V''| < \log n$ steps, such that v_i is smaller than v, v is smaller than v_j, and moreover $i < j < i + 4c\log n$. Since i and j belong to some initial values $r = r_i, r_j$ of Lemma 3 with $|r_i - i| < 2c\log n$ and $|r_j - j| < 2c\log n$, we can immediately conclude that v is larger than at least $i - 4c\log n$ vertices of V', and smaller than at least $n' - j - 4c\log n$ vertices of V'. Thus, with at most $12c\log n$ further questions we can detect all orientations between v and V' not known so far.

Since the number of steps involving v is less than $13c\log n$, the total number of questions required for $G_{n,p}$ does not exceed $O(n\log n)$. □

3 Sparse random graphs

In this section we prove Theorem 2.

Given a graph $G = (V, E)$, let the *random strategy* be the following strategy of player A: pick a random permutation $e_1, e_2, \dots, e_m$ of the edges of G and ask for the orientation of the edges in this order, where the orientation of the edge e_i is probed if and only if it does not follow from the orientations of the edges $e_1, \dots, e_{i-1}$ (and the assumption that the orientation is acyclic). We claim that for every edge probability p, if player A applies this strategy on the random graph $G = G_{n,p}$, then almost surely he will not have to ask more than $O(n \log^3 n)$ questions *even if he tells the order* in which he is going to ask the questions to the Strategist already at the beginning. This clearly implies the assertion of Theorem 2. The advantage in considering this variation of the game is that since the first player A announces his full strategy already at the beginning, the second player S does not have to decide step by step; instead, he can create his strategy at once, by choosing an acyclic orientation of G.

Therefore, the version of the game we consider now is as follows. Player A chooses a random permutation $e_1, e_2, \ldots$ of the edges of $G = G_{n,p}$ and reports it to the Strategist. The Strategist next chooses a linear order on the vertices of G and orients its edges according to this order (by orienting each edge from its smaller end to its larger end). The value of the game is the number of edges e_i in the oriented graph G that do not lie in the transitive closure of the oriented edges $e_1, \ldots, e_{i-1}$, as this is the number of questions A will actually have to ask. Therefore, our objective is to prove the following.

Proposition 4 *For any edge probability* p *and for a random ordering* $e_1, e_2, \ldots$ *of the set of edges of the random graph* $G_{n,p}$*, the following holds almost surely. For every linear order of the vertices of* G *and for the associated acyclic orientation of* G*, the number of oriented edges* e_i *that do not lie in the transitive closure of the oriented edges* $e_1, \ldots, e_{i-1}$ *does not exceed* $O(n \log^3 n)$.

Notice that the subgraph of $G_{n,p} = (V, E)$ consisting of its first i randomly chosen edges $e_1, \ldots, e_i$—denoted by G_i—is simply a random graph with i edges and n vertices. This fact plays a crucial role in our proof. It is worth noting that in view of this fact it suffices to prove the above proposition for the case $p = 1$, i.e., for the case that G is the complete graph. However, since this does not simplify the argument, we consider the general case $G = G_{n,p}$.

The proof relies on some of the ideas applied in the study of parallel comparison algorithms for approximation problems (see [2], [9], [3], [4]). In particular, we need the following known result implicit in [2] (cf. [9], [3]).

Lemma 5 *There exists an absolute constant* $b > 0$ *with the following property. Let* G *be a graph on* n *vertices in which there is at least one edge between any two disjoint sets of* q *vertices each. Then, the number of edges in the transitive closure of any acyclic orientation of* G *is at least* $\binom{n}{2} - bnq \log n$. □

The next lemma can be proved by a straightforward calculation which we omit.

Lemma 6 *There exists an absolute constant* c *so that for every* i*,* $n \log n \leq i \leq \binom{n}{2}$*, if* G_i *is a random graph with* n *vertices and* i *edges, then the probability that* G_i *has at least one edge between any two disjoint sets of* $(cn^2 \log n)/i$ *vertices each is at least* $1 - 1/n^{\log n}$. □

Proof of Proposition 4. Throughout the proof we assume, whenever this is needed, that n is sufficiently large. To simplify the presentation, we make no attempt to optimize the various multiplicative constants appearing here. Recall that for each admissible $i \geq 1$, G_i denotes the subgraph of $G = G_{n,p}$ consisting of the edges $e_1, \ldots, e_i$. By Lemma 6, and since each G_i is a random graph with i edges, the following event denoted by $\mathbf{E}$ occurs almost surely: for every $i \geq n \log n$ there is an edge of G_i between any two disjoint sets of $(cn^2 \log n)/i$ vertices each.

Fix a linear order L on the vertices of G, and let $\mathbf{E}_L$ denote the event that the number of edges e_i that do not lie in the transitive closure of the edges $e_1, \ldots e_{i-1}$ once these are oriented according to L exceeds $32bcn \log^3 n$, where b and c are the constants from Lemmas 5 and 6, respectively. We next show that for each fixed L, the conditional probability $Prob[\mathbf{E}_L|\mathbf{E}]$ is much smaller than $1/n!$. To do so, let us split the choice of the edges of $G_{n,p}$ and the random permutation on them into phases as follows. For each power of 2, i.e. $2^j \geq 1$, phase j consists of the choice of the edges e_r for all $2^j \leq r < 2^{j+1}$ of G (assuming G has at least that many edges). An equivalent, more precise, description of the random procedure of choosing the edges e_i in the various phases is as follows. First choose the number m of edges of G according to a binomial distribution: $Prob[m = s] = \binom{N}{s} p^s (1-p)^{N-s}$, where $N = \binom{n}{2}$. Next, starting with $j = 0$, in phase j choose 2^j edges at random among the ones not chosen so far, as long as $2^{j+1} - 1 \leq m$. In the last phase, the one corresponding to the largest j for which $2^j \leq m$, we choose only $m - 2^j + 1$ random edges. Let $\mathbf{E}_{L,j}$ denote the event that during phase j the number of edges e_r that do not lie in the transitive closure of the first $2^j - 1$ oriented edges exceeds $16bcn \log^2 n$. Since $\mathbf{E}_L$ is contained in the union $\cup_{j \geq 0} \mathbf{E}_{L,j}$ (as the number of phases is less than $2 \log n$), we have

$$Prob[\mathbf{E}_L|\mathbf{E}] \leq \sum_{j \geq 0} Prob[\mathbf{E}_{L,j}|\mathbf{E}].$$

If $2^j \leq 16bcn \log^2 n$, then clearly $Prob[\mathbf{E}_{L,j}|\mathbf{E}] = 0$. For any larger j, observe that

$$Prob[\mathbf{E}_{L,j}|\mathbf{E}] = \frac{Prob[\mathbf{E}_{L,j} \wedge \mathbf{E}]}{Prob[\mathbf{E}]} \leq 2Prob[\mathbf{E}_{L,j} \wedge \mathbf{E}].$$

(Here we used the fact that $Prob[\mathbf{E}] \geq 1/2$; in fact this probability is $1 - o(1)$.) However, if $\mathbf{E}$ happens then, by Lemma 5, the transitive closure of the graph G_{2^j-1} (oriented according to L) contains at least $\binom{n}{2} - \frac{bcn^3 \log^2 n}{2^j - 1}$ edges. If $2^j - 1 \geq n^2/16$, then certainly the event $\mathbf{E}_{L,j}$ will not occur, as the total number of edges which are not in the transitive closure we consider is at most $16bcn \log^2 n$. Otherwise, in phase j we are choosing 2^j $(\leq n^2/16 + 1)$ edges at random among the $\binom{n}{2} - 2^j + 1 \geq (1 + o(1)) \frac{7}{16} n^2$ remaining ones, and the number of edges among those which are not in the transitive closure of G_{2^j-1} is at most $\frac{bcn^3 \log^2 n}{2^j - 1}$, i.e., a fraction of at most

$$(1 + o(1)) \frac{16bcn \log^2 n}{7(2^j - 1)} < 3bcn \log^2 n / 2^j$$

of the remaining edges (here we assumed that n is large enough). Therefore, the expected number of edges chosen in the j^{th} phase which are not in the transitive closure of G_{2j-1} is smaller than $3bcn\log^2 n$. By standard estimates (see, e.g., [5], Appendix A, Theorem A.12) it follows that the probability that more than $16bcn\log^2 n$ such edges are chosen (i.e, that $\mathbf{E}_{L,j}$ happens) is at most $exp\{-\Omega(n\log^2 n)\}$. This bounds $Prob[\mathbf{E}_{L,j} \wedge \mathbf{E}]$ and hence $Prob[\mathbf{E}_{L,j}|\mathbf{E}]$ as well, and implies that for every fixed L, $Prob[\mathbf{E}_L|\mathbf{E}] \le 2\log n\ exp\{-\Omega(n\log^2 n)\}$.

To complete the proof of the proposition, observe now that the probability that there exists a linear order L for which there are more than $32bcn\log^3 n$ edges e_i that do not lie in the transitive closure of the previous edges is at most

$$Prob[\overline{\mathbf{E}}] + \sum_L Prob[\mathbf{E}_L|\mathbf{E}] \cdot Prob[\mathbf{E}] \le o(1) + 2n!\log n\ exp\{-\Omega(n\log^2 n)\},$$

which tends to zero as n tends to infinity. This completes the proof of Proposition 4, and implies the assertion of Theorem 2. □

4 Exhaustive graphs

Trivially, any (acyclic) orientation of a graph $G = (V, E)$ can be identified by $|E|$ questions. Call G *exhaustive* if it admits nothing better than this trivial algorithm, i.e., if $c(G) = |E|$. We do not know too much about the structure of exhaustive graphs. It is observed in [1] that every bipartite graph is exhaustive, and it is also shown there that exhaustive graphs on n vertices have at most $\frac{1}{4}n^2$ edges (for all $n \ge 6$). Using arguments similar to those used in the proof of Theorem 2 we can show that a random graph with n vertices and more than $n\log n\log\log n$ edges is almost surely non-exhaustive. Similar techniques can be used to show that there are non-exhaustive graphs of arbitrarily high girth. A couple of small non-exhaustive graphs are mentioned in [1]. The next proposition exhibits a further explicit example and answers a question raised in [1], where the authors wonder if there are line graphs of triangle-free cubic graphs which are non-exhaustive.

Proposition 7 *The line graph $L(K_{3,3})$ of the complete bipartite graph $K_{3,3}$ is non-exhaustive.*

Proof. If three vertices x, y, z induce a triangle in an exhaustive graph and the orientation of precisely one edge, say $x \to z$, is known, then the next answer concerning the edge xy or yz is determined, namely if xy (yz) is probed next then the answer must be $x \to y$ ($y \to z$), for otherwise the orientation of the third edge of the triangle were determined by the other two. For such situations we shall use the shorthand "$x \to z$ *forces* $x \to y$" or "$x \to z$ forces $y \to z$" which will also mean that the next edge asked is xy or yz, respectively.

Suppose now on the contrary that $L(K_{3,3})$ is exhaustive. Assuming that the vertex classes of $K_{3,3}$ are $\{x_1, x_2, x_3\}$ and $\{y_1, y_2, y_3\}$, we denote by v_{ij} the vertex of $L(K_{3,3})$ that represents the edge x_iy_j; hence, v_{ij} and $v_{i'j'}$ are adjacent if and only if $i = i'$ or $j = j'$. At the beginning we ask about the orientations of $v_{11}v_{12}$ and $v_{13}v_{23}$. By symmetry, we may assume without loss of generality that these two orientations are $v_{11} \rightarrow v_{12}$ and $v_{13} \rightarrow v_{23}$. Then we ask about $v_{21}v_{31}$ and prove that either answer will allow us to save at least one question.

Suppose first $v_{21} \rightarrow v_{31}$. Then $v_{21} \rightarrow v_{31}$ forces $v_{21} \rightarrow v_{11}$ and $v_{11} \rightarrow v_{12}$ forces $v_{11} \rightarrow v_{13}$, therefore the directed path $v_{21} \rightarrow v_{11} \rightarrow v_{13} \rightarrow v_{23}$ determines the orientation of $v_{21} \rightarrow v_{23}$ and this question need not be asked. Hence, suppose $v_{31} \rightarrow v_{21}$. Then $v_{31} \rightarrow v_{21}$ forces $v_{31} \rightarrow v_{11}$, $v_{11} \rightarrow v_{12}$ forces $v_{11} \rightarrow v_{13}$, and $v_{13} \rightarrow v_{23}$ forces $v_{13} \rightarrow v_{33}$. Thus, the directed path $v_{31} \rightarrow v_{11} \rightarrow v_{13} \rightarrow v_{33}$ determines the orientation of $v_{31} \rightarrow v_{33}$. $\square$

We note that apart from the density-type results, so far the non-exhaustiveness of particular graphs has been proved by *ad hoc* arguments. It would be interesting to know more about the structural reasons that make a graph non-exhaustive.

References

[1] M. Aigner, E. Triesch and Zs. Tuza, *Searching for acyclic orientations of graphs*, to appear.

[2] M. Ajtai, J. Komlós, W.L. Steiger and E. Szemerédi, *Almost sorting in one round*, Advances in Computing Research, Vol. 5, 1989, JAI Press, 117-126.

[3] N. Alon and Y. Azar, *Sorting, approximate sorting and searching in rounds*, SIAM J. Discrete Math. 1 (1988), 269-280.

[4] N. Alon and Y. Azar, *Parallel comparison algorithms for approximation problems*, Proc. 29^{th} IEEE FOCS, Yorktown Heights, NY, 1988, 194-203. Also: Combinatorica 11 (1991), 97-122.

[5] N. Alon and J. H. Spencer, *"The Probabilistic Method"*, Wiley, 1991.

[6] B. Bollobás, *"Random Graphs"*, Academic Press, 1985.

[7] N. Kahale and L. Schulman, *Bounds on the chromatic polynomial and on the number of acyclic orientations of a graph*, to appear.

[8] U. Manber and M. Tompa, *The effect of the number of Hamiltonian paths on the complexity of a vertex coloring problem*, Proc. 25^{th} IEEE FOCS, Singer Island, Florida 1984, 220-227.

[9] N. Pippenger, *Sorting and selecting in rounds*, SIAM J. Comput. 6 (1987), 1032-1038.